Physical Exercise and Reduced Risk of Breast Cancer in Young Women.

Bernstein L, Henderson BE, Hanisch R, Sullivan-Halley J, K. Ross RK.

Journal of the National Cancer Institute, vol 86, September 21, 1994.

STUDY DESIGN AND CONDUCT

Study Aim

The major objective of the study was to determine whether young women (aged 40 and younger) who regularly participated in physical activities during their reproductive years had a reduced risk of breast cancer. The objective was to be answered using a case-control study.

The rationale is that ovarian hormones may be a determinant of breast cancer risk, and physical activity seems to modify ovarian hormonal production, thus decreasing breast cancer risk.

There are no other objectives mentioned, including secondary endpoints. It is noted that the authors do analyse other endpoints: examples include determining whether there was a difference in results for patients with invasive cancer alone, and whether there was any difference between nulliparous and parous women. These secondary endpoints should have been stated prospectively; if they were not then they may be the results of a data dredging exercise (thus introducing bias). This aspect requires clarification by the authors.

Sample size and power

There are no calculations of required sample size or power. According to Schlesselman, "the number of subjects to be selected for study of a specific disease-exposure relationship is a fundamental consideration in planning a case-control investigation" (1982. p.144).

Cases

All white female residents in LA County diagnosed with in-situ or invasive cancer and registered in the Cancer Surveillance Program between 1/7/83 and 1/1/89 were eligible, provided they were 40 yo or younger at diagnosis and born in the USA , Canada or Europe. This is a fairly general young population, although only whites were included who were born in the above countries. It is not clear why birthplace was restricted. The study was limited to white women "because of the small representation of individual minority groups among young breast

Cases were identified using a cancer registry. The quality of the results are only as good as the quality of the records. Problems may occur due to incomplete data, data entry error, or patients from the area being diagnosed or seeking cancer treatment elsewhere. Physicians may differ in the criteria they apply and in their skills for diagnosing certain diseases, diligence in the recording of results, and false positive diagnoses may occur (Colton, 1974, p.280). Thus quality control of a case-control study is important: for example researchers might ask for the pathology of those cases identified to be reviewed. No quality control measures are mentioned, and the authors should attempt to describe any measures instituted.

A total of 969 patients were eligible. Of these, 20 had died, 54 were refused contact by the treating physician, 111 refused, 12 had moved and 21 were lost to follow-up. Thus interviews were performed on 744 patients. Of the 744

The fact that interviews could not be performed in all patients is one of the problems of an observational study like this: no matter how much effort the researchers go to, it would be virtually impossible to be able to interview all of the identified patients. This lack of information introduces the possibility of bias, for example the fact that patients refused to be interviewed may mean that they are less healthy (and may have exercised less earlier in life). The 199 patients excluded due to differences in questionnaires may have had different characteristics to those cases interviewed. Such bias decreases the ability to extrapolate results to the general population. In an attempt to overcome the possible differences in the group of the 545 analysed and the other eligible patients not assessed, a comparison could be sought to show that participants do not have atypical features that may affect the results (8). The authors should state whether this was performed, and whether reason for refusal was obtained.

Controls

Controls were selected and individually matched to the 744 interviewed cases.

Controls were selected using a fairly random method. The neighbourhood that the control came from was that of the case at her time of diagnosis: this was used to match the case and control in terms of neighbourhood of residence. This method does not allow for the fact that either the case or the control may have spent their entire previous lives in another place. Also, since the date of diagnosis could be anywhere between about 5 and 10 years prior to finding the control, the neighbourhood may have changed considerably with regards to variables influencing breast cancer incidence or exercise patterns. The authors should note whether there was any attempt

For 592 patients, the first eligible control subject participated. Unfortunately, as is usually the case with such studies, a certain number of potential controls will refuse to participate. For 124 case patients, 1 eligible person refused. For the remaining 28, the number of refusals ranged from 2 to 6. Once again, we must ask ourselves

The issue of exclusion of non-whites from the study is of importance, as previously discussed.

One "neighbourhood control" patient was individually matched to each of the 744 interviewed case patients on birthdate (within 36 months), race (white), and parity (nulliparous versus parous). Control subjects were also restricted as to birthplace. Other potential confounding variables are compared between the two groups (to be further discussed).

Questionnaire

In-person interviews were conducted by the same female nurse-interviewer, and a complete reproductive, contraceptive and physical exercise history was obtained, as well as other relevant information.

The wording of the questions must be such as to avoid questionnaire bias (Bland, 1995, p.40-41). The way in which the interviewer asks questions may influence the answers, thus the interviewer should be blind (8). There is no indication of the degree of blindness in this study, and this should be addressed by the authors.

If the cases are aware of the risk factor (ie lack of exercise), then they may overreport the risk factor in themselves (8). This is called recall bias, and efforts should be made to evaluate and minimize the effect of recall bias (Altman, 1991, p. 94). Another possible bias is detection bias: women who exercise may be more health conscious and thus be more likely to examine their breasts themselves or by a physician.

Another area that requires further elaboration by the authors is the reason for changing the questionnaire after the first 199 patients. Was the information obtained inadequate for the intended research due to poor questionnaire, was the method of calculating amount of exercise changed after the trial had commenced, or were the results obtained in these patients not what the investigators were hoping for? Was this change of questionnaire and exclusion of these patients from analysis allowed for in the study protocol, or was it determined while the study was underway? Was an attempt made to determine whether the 199 case-control pairs excluded were atypical compared to those included in the study. These questions must be addressed by the authors due to the large proportion of case-controls excluded for this reason (almost 27% of those interviewed).

Questionnaires may be inaccurate in obtaining a true picture of a person's exercise habits. A retrospective study relies almost entirely on a person's memory: in this case details over the last decades including childhood memories. It is unlikely that exact details can be obtained for all case-control pairs. The authors state that complete histories were obtained in all case-control pairs interviewed with the second questionnaire.

The validity of the measures of physical activity was not addressed by the authors (as pointed out in Brinton's editorial (4). Walking for exercise may involve a slow hour long stroll around the block, whereas an hour of squash is far more demanding. Occupations involving exercise, such as a postwoman or bicycle courier are not included as exercise, despite the fact that after a day delivering mail, few people would feel like going for a walk for exercise purposes. In their own discussion, the authors mention a study that found a decreased risk of breast cancer in physical education teachers compared to language teachers, thus it is interesting that type of occupation was not considered.

The questionnaire does not take into account several important factors which are not subsequently assessed. Family history does not take into account second degree relatives, number of relatives with breast cancer, diagnosis of relatives at age less than 40, male relatives with breast cancer, bilaterality of cancer, or family histories of ovarian cancer or sarcoma. All of these factors contribute to risk of breast cancer. Quetelet's index is determined at the reference date, and does not consider changes in weight in the woman's past. Other potential risk factors (as mentioned by Brinton) such as diet, alcohol intake, vitamin intake, and occupations (such as electrical line workers or exposure to organochlorines) have not been considered: uneven distributions between cases and controls may affect the results.

ANALYSIS AND RESULTS

Patient characteristics

Potential confounding factors in terms of patient attributes are listed in Table 1. The distribution of these characteristics seems fairly even between the cases and controls. These factors were considered in the logistic regression analyses. The authors do not consider other potential confounders such as occupation, family history and diet, as previously discussed.

Statistical analysis

The statistical methods used are listed and referenced in the methods section and are appropriate for this study.

One of the problems with the statistical analysis is the multiple testing performed, particularly when only one objective has been stated. It is certainly not unreasonable for the authors to, for example, re-analyse only case-control pairs where the case was diagnosed with invasive breast cancer: however the desire to do so should have been stated in the trial protocol (ie in advance). This avoids the retrospective "dredging" of data. The authors should state at the outset precisely what the trial intends to investigate, in terms of major and (a few) minor objectives.

Another problem which has been previously discussed, is the failure to include a large proportion of case-control women in the analysis, for reasons unexplained. We have no way of knowing whether excluded women differ in any way from analysed women: thus extrapolation of results is questionable. The authors need to clearly state why women were not included (in particular the 199 pairs who were initially questioned), and whether they may have differed from the women analysed.

For exercise factors, approximate quartiles were created based on the distribution of control subjects who reported any activity. The intent to do this should have been prospectively stated, in order to avoid bias. Alternatively, the divisions could have been set before the trial: approximate quartiles could have been estimated by a small survey, or divisions could be based on whole hours of exercise.

Confidence intervals have been well utilised, however a perusal of the results show that many of the 95% confidence limits for the odds ratios are on either side of 1.0. This indicates that the odds ratio in such instances could actually be higher than 1.0 (ie an increased risk of breast cancer with certain levels of exercise). The authors seem to concentrate on presenting the odds ratios that have confidence limits under 1 (those for the highest levels of exercise), without acknowledging it.

In the results of exercise within the 10 years after menarche, all of the confidence limits straddle 1. The authors state that the addition of months of activity after this 10-year time period to the multivariate logistic regression model significantly improved the fit of the model. The authors then use this finding to conclude that "the protective effect of exercise on breast cancer risk in the women we studied was more pronounced when we considered the woman's complete exercise history than when we restricted our analyses to the activities during the 10 years after menarche". The division of exercise into during and after 10 years after menopause seems to have been arbitrary and retrospective, once again introducing bias. If this was contemplated at the outset of the study then it should be stated as a secondary endpoint to be analysed. It is also noted that in the "within the 10 years after menarche" analysis, 3 cases were excluded since there was less than 10 years time between menarche and reference date (a small number overall).

The odds ratios were tested for linear trend across categories, with the results found to be significant for the overall analysis. However there is a non-linear variation in these odds ratios for hours worked which indicates that the linear trend only explains some of the exercise effect. This has not been commented on by the authors.

Several other analyses seem to have been performed retrospectively, including employment status at reference date (which it is acknowledged cannot possibly represent lifetime employment history), analysis by parity (and on finding a difference, further analysing by marriage history), invasive breast cancer (for this all analyses were apparently repeated), and other variables (including 3 variables acknowledged as being "created" by the authors). For many of these analyses, actual results are not shown. Multiple analyses such as these must be interpreted cautiously, especially if determined retrospectively in a data dredging exercise. The authors should clearly state in the objectives which of these analyses were contemplated prospectively.

According to Altman (1991, p.94), any variable used for matching cannot be investigated as a possible risk factor for the outcome. Thus, since cases have been individually matched with controls with regards to parity (nulliparous versus parous), it is not possible to find an association between breast cancer and parity if there is one. The authors determine the relative odds of breast cancer in most active versus inactive women among parous and among nulliparous women, and then compare the two relative odds in the discussion. This comparison, as pointed out by Schlesselman (1982, p. 120), would not seem valid, and the authors should justify such an analysis, or exclude it.

Presentation

The overall presentation of the results is good.

TRIAL CONCLUSIONS

The authors arrive at several conclusions in the discussion.

Continued participation in physical exercise activities during a woman's reproductive years, particularly among women of proven fertility, was found to have a striking protective effect on breast cancer risk, and that this is consistent with their hypothesis that habitual exercise should reduce the risk of breast cancer by altering menstrual function.

The "data strongly suggests that continued participation in a physical exercise regimen can markedly reduce the risk of breast cancer in premenopausal women and emphasize the importance of beginning an exercise regimen early in life and maintaining it during adulthood".

The "results strongly support the need for educational policies that require participation in physical education classes and that encourage lifelong participation in exercise programs".

These statements are implying that a causal link has been established by the study between lack of exercise and breast cancer, which is not true: an association has been found. "Observational studies cannot do more than suggest possible causal links" (Altman, 1993, p.96). This association seems to be mainly for the higher levels of exercise, and only relates to risk of breast cancer under 40 years of age (which account for about 20% of all breast cancers. If it can be proven that exercise when a woman is young does decrease breast cancer risk before the age of 40, is it possible that exercise may only delay the risk of breast cancer until a later age, and is post-menopausal breast cancer a different disease altogether? The authors also infer that the results can be extrapolated to the population as a whole. This is not the case: results could only be extrapolated to white females under 40 who have not previously had breast cancer, who were born in the USA, Canada or Europe. The generalizability of the results is also under question due to the analysis being based on 545 of 969 eligible patients, and the substantial refusal rate of potential controls.

The protective effect of exercise did not appear to be mediated by an effect on body mass, and measures of childbearing and childrearing did not modify the effects of exercise.

The authors give no explanation of why these results should differ from those of other studies. For example, they could comment on the fact that a single obesity index does not necessarily represent the woman's weight history over the previous decades.

The data suggest that women who maintain an activity level of 1-3 hours/week could reduce their risk of breast cancer by about 30% relative to inactive women, and those that maintain an activity level of at least 4 hours per week could reduce their risk by more than 50%.

It is noted however that the quartiles analysed did not correspond to 1-3 hours and 4 or more hours: ie the analysis of 1-3 hours and 4 or more hours was not performed. Also the authors give no confidence intervals for this statement: in particular they again fail to comment on the confidence limits obtained for "medium level" activities which span an odds ratio of 1.

The inability to demonstrate a "dose-response" effect among nulliparous women (ie using the test for linear trend) may be due to the inclusion of a subset of women who have an underlying ovarian problem.

This is based on results obtained by analysing marital status, assuming that never married women were less likely to be nulliparous because of an underlying ovarian problem than ever married women. The conclusion cannot be sustained by such an assumption, as it can't possibly be validated. Retrospective analysis of this type to try to explain results should not be performed, and if they are, results should be viewed cautiously. The validity of the initial analysis itself is in question, as previously discussed, since the cases and controls were matched for parity (nulliparous versus parous).

The protective effect of exercise on breast cancer risk in the women studied was more pronounced when the woman's complete exercise history was considered, rather than restricting analyses to the activities during the 10 years after menarche. This emphasizes the importance of establishing life-long patterns of physical activity.

This result was interesting, and the desire to analyse the data in this way should have been stated in the trial protocol to lend more weight to results. The authors again do not comment on the confidence limits. They also do not consider the possibility that if there is a benefit conferred by exercise, that it may be far more important for exercise after the 10 year post-menarche period (and that life-long exercise may not be necessary).

RECOMMENDATIONS

I believe that the study is of great interest despite the flaws discussed, but I would suggest the authors review several aspects before its publication:

>A clear statement of the primary and secondary outcomes that had been prospectively determined and were to be measured in subsequent analyses.

>Discuss whether any sample size or power calculations were performed.

>Outline any quality control measures that were undertaken in this study.

>Explain more clearly the reason why non-whites were excluded from the study, and was the decision to exclude them made only after the study commenced.

>Indication of reason for refusal of cases and controls to participate in the study, and whether those that refused had atypical characteristics compared to those that agreed to participate.

>Comment on why factors such as type of occupation, diet and full family history of breast cancer were not considered important.

>Discuss how valid the authors believe the measures of physical activity are in this group of women for this study design.

>Acknowledge whether the interviewer was blind, and whether the interviewees were blind as to the purpose of the study.

>Explanation as to the reason the questionnaire was altered during the study, why the first 199 case-controls were

excluded, and whether there was any difference between these case-controls compared to those included in the analysis.

>Comment on the findings of odds ratios with confidence intervals on either side of 1.0, and discuss the finding of non-linear variation in these odds ratios for hours worked.

>Consider whether the results obtained suggest that any benefit of exercise, if proved, may only be with higher levels of exercise.

>Comment on whether it is valid to attempt to find an association between breast cancer and parity, given that parity was a variable used for matching.

>Comment on whether the authors believe that causality has been established, particularly in view of the weakness of the association found.

>Discuss what population their results may be extrapolated to in view of the study design and weaknesses